2019

SKELLAM PROCESS MODELING FOR FINANCIAL HIGH-FREQUENCY DATA

Tingfang Lee
University of Rhode Island, tingfanglee@uri.edu

SKELLAM PROCESS MODELING FOR FINANCIAL HIGH-FREQUENCY DATA

BY

TINGFANG LEE

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

STATISTICS

UNIVERSITY OF RHODE ISLAND

2019

MASTER OF SCIENCE THESIS

OF

TINGFANG LEE

APPROVED:

Thesis Committee:

Major Professor Gavino Puggioni

Jing Wu

Jing Wu

Shingo Goto

Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2019

ABSTRACT

High-frequency data are observations collected at fine time scale. Such data largely incorporates pricing and transactions, of which institutional rules prevent from drastically rising or falling within a short period of time. This results in data changes based on the measure of one tick, a measure of the minimum upward or downward movement in the price of a security. The discreteness brings that the observations are in $\mathbb{Z}$. A Skellam distribution has a unique property that returns values in $\mathbb{Z}$.

We are interested in studying the Skellam process where the time-dependent intensities are Gaussian process. Such doubly stochastic Poisson process, also known as Cox process, is a point process which is a generalization of a Poisson process. We then investigate if this Skellam model provide better fit to high frequency financial data and how Gaussian process can capture the market microstructure.

ACKNOWLEDGMENTS

I would first like to thank my thesis advisor Dr. Gavino Puggioni. The door to Prof. Puggioni office was always open whenever I ran into a trouble spot or had a question about my research or writing. He has been supportive of my career goals and who worked actively to provide me with the protected academic time to pursue those goals.

I am grateful to all of those with whom I have had the pleasure to work during this and other related projects. Each of the members of my thesis committee has provided me extensive personal and professional guidance.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

High-frequency data consists of observations collected at a fine time scale. Each single observation (transaction, quote, price movement, etc) is characterized by one logical unit of information. Such data largely incorporates pricing and transactions which institutional rules prevent from drastically rising or falling within a short period of time. This results in data changes based on the measure of one tick, a measure of the minimum upward or downward movement in the price of a security. For instance, an uptick indicates a trade where the transaction has occurred at a price higher than the previous transaction, and a downtick indicates a transaction that has occurred at a lower price. The development in computer technology and data recording and storage have made these datasets increasingly accessible to researchers and have driven the data frequency to its limit for some financial markets. For equity markets, the Trades and Quotes (TAQ) database of the New York Stock Exchange (NYSE) contains all recorded trades and quotes on NYSE, AMEX, NASDAQ, and the regional exchanges from 1992 to present. The Berkeley Options Data Base recorded similar data for options markets from 1976 to 1996.

These high-frequency financial datasets have been widely used to study numerous market microstructure related questions, including price discovery, competition among related markets, strategic behavior of market participants, and modeling of real time market dynamics. Moreover, high frequency data is also useful for studying statistical properties, volatility, in particular, of asset returns at lower frequencies. There are many methods of modeling market microstructures with different procedures and underlying assumptions. For instance, the GARCH model is one

popular statistical instrument that captures the dynamics of asset return volatility. The models are standard tools for describing and forecasting time-varying volatility. Luca (2006) [10] discussed GARCH modeling on high frequency data to predict short term volatility and understand how the use of high frequency data could be a support for the analysis carried out on daily data. Visser (2011) [7] developed a quasi maximum likelihood estimator that uses volatility proxies to estimate the parameters of a discrete-time, daily GARCH model. Chavez-Demoulin and McGill (2012) [2] were interested in events that lead to extreme losses and concentrated on the left tail of the distribution function of the log-returns. They modeled excesses of high frequency financial time series via a Hawkes process. Some other excellent reviews on the use of high frequency financial datasets in financial econometrics can be found in [1, 3, 4, 5, 6].

Due to the discreteness of the observations in high frequency data, the recent literature includes models that use the difference of two Poisson processes, and incorporate a Skellam distribution for forward prices. Kerss et al. (2014) [9] defined the fractional Skellam processes via time changes in ordinary Skellam processes. Although the exponential distribution of inter-arrival times in models is not always supported by data, this fractional Skellam model incorporates the Mittag-Leffler distribution of inter-arrival times and overcomes that limitation. Koopman et al. (2017) [8] developed the dynamic Skellam model that is empirically relevant for the discrete time series of tick-by-tick price changes. They capture the intraday seasonal patterns by including a spline function over the time period of a single day. In addition, they also allowed for autoregressive intraday stochastic volatility dynamics to capture any remaining volatility dynamics over the course of the trading day.

These models present possible efficient estimators and predictors for volatility,

time-varying correlation structures, trading volume, bid-ask spreads, depth, trading costs, and liquidity risks. The estimates from these models provide valuable input regarding trading algorithms and academic research. Modeling and understanding the interaction between liquidity supply and demand is fundamental to characterizing various aspects of the trading process and investors behavior. In this study, we build a new statistical model that has more flexibility to capture the market microstructure in high frequency data. The model is built on the Skellam process, the difference of two Poisson processes. Such random variables take values in $\mathbb{Z}$ and hence the tick-by-tick price change can be treated as the Skellam distribution. We then use Gaussian processes to model the intensities of the Poisson processes in the logarithmic scale the intensity functions of the two Poisson processes whose difference yields the Skellam process. This semiparametric approach allows us to model data collected at not-equally spaced intervals, while accounting for nonlinearities in their time evolution. The rest of Chapter 1 will briefly introduce the definition of the Skellam process and the Gaussian process, then illustrate the main objective of the thesis.

1.1 The Skellam distribution

Definition 1.1. Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time with known intensity.

The probability mass function of a Poisson distribution with given intensity $\lambda > 0$ is

$$p(k; \lambda) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

The positive real number λ is equal to the expected value and also to its variance.

Definition 1.2. The Skellam distribution [12],

$$S(\lambda_1, \lambda_2) = \text{Poisson}(\lambda_1) - \text{Poisson}(\lambda_2),$$

is the discrete probability distribution of the difference of two independent Poisson distributions. Each Poisson distributed with respective intensities λ_1 and λ_2.

The probability mass function of $S(\lambda_1, \lambda_2)$ is

$$s(k; \lambda_1, \lambda_2) = e^{-(\lambda_1+\lambda_2)}\Big(\frac{\lambda_1}{\lambda_2}\Big)^{k/2} I_k(2\sqrt{\lambda_1\lambda_2})$$

where $k \in \mathbb{Z}$ and $I_k(z) = \displaystyle\sum_{m=0}^{\infty} \frac{1}{m!\Gamma(m+k+1)}\Big(\frac{z}{2}\Big)^{2m+k}$ is the modified Bessel function of the first kind. The expected value is $\lambda_1 - \lambda_2$, while the variance is $\lambda_1 + \lambda_2$. Moreover,

$$s(k; \lambda_1, \lambda_2) = s(-k; \lambda_2, \lambda_1)$$

such that the Skellam distribution is symmetric when $\lambda_1 = \lambda_2$, right-skewed when $\lambda_1 > \lambda_2$, and left-skewed when $\lambda_1 < \lambda_2$.

1.2 The Skellam Process

Definition 1.3. A Skellam process is defined as

$$S_t = S(\lambda_{1,t}, \lambda_{2,t}) = \text{Poisson}(\lambda_{1,t}) - \text{Poisson}(\lambda_{2,t})$$

where $t \geq 0$ and $P(\lambda_{1,t})$ and $P(\lambda_{2,t})$ are two independent Poisson processes with intensities $\lambda_{1,t} \geq 0$ and $\lambda_{2,t} \geq 0$.

1.3 The Gaussian Process

A Gaussian process (GP) is a stochastic process, which provides a powerful tool for probabilistic inference directly on distributions over functions and which has gained much attention in recent years. While being well known and extensively used in the statistical literature (for instance in spatial and spatio-temporal

models) Gaussian processes have received considerable attention in the machine learning literature in several nonlinear regression and classification problems [11]. It offers a flexible non-parametric Bayesian framework for estimating latent functions from data.

Definition 1.4. A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution.

A GP is completely specified by its mean function and covariance function which determine the smoothness and variability of the functions. Given input vectors $\mathbf{x}$ and $\mathbf{x}'$, we define mean function $m(\mathbf{x})$ and the covariance function $k(\mathbf{x}, \mathbf{x}')$ of a real process $f(\mathbf{x})$ as

$$m(\mathbf{x}) = E[f(\mathbf{x})]$$
$$k(\mathbf{x}, \mathbf{x}') = E[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$$

and will write the Gaussian process as

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')).$$

1.4 Main goals

Market microstructure is the study of financial markets and how they operate. It is one of the fastest growing fields of financial research due to rapid development of algorithm and electronic trading. Such high speed evolution of financial markets results in tremendously amount of trading venues and exchanges. The major thrust of market microstructure research examines the ways in which the working processes of a market affect determinants of transaction costs, prices, quotes, volume, and trading behavior. We introduce a new modeling extension for the Skellam process, which includes a non parametric GP specification for the log intensity functions. We test the new model with simulated data. For illustrative purposes we show also the analysis of some real data, collected at moderate frequency.

Chapter 2 describes the Skellam process model with Gaussian process priors. We then introduce the Laplace approximation and MCMC methods to approach the conditional posterior with Skellam likelihood.

Next, we use simulated data to investigate the performance of the Skellam model in terms of prediction in Chapter 3. In addition, we compare the prediction between Gaussian likelihood and Skellam likelihood.

Chapter 4 presents the empirical results related to model fit and the forecasting performance on three high frequency datasets from S&P 500 index, NASDAQ, and IBM stock trading data in 1-min frequency.

In the last chapter, we will discuss the results and direction of future research.

List of References

[1] Torben G. Andersen, *Some Reflections on Analysis of High-Frequency Data,* Journal of Business and Economic Statistics, Volume 18, No. 2, 2000, Pages 146-153.

[2] Valerie Chavez-Demoulin, James A. McGill, *High-frequency financial data modeling using Hawkes processes,* Journal of Banking and Finance, Volume 36, Issue 12, December 2012, Pages 3415-3426

[3] John Y. Campbell, Andrew W. Lo, A.Craig MacKinlay, *The Econometrics of Financial Markets,* Princeton University Press, 1997.

[4] Eric Ghysels, *Some Econometric Recipes for High-Frequency Data Cooking,* Journal of Business and Economic Statistics,Volume 18, No. 2, 2000, Pages 154-163.

[5] Ramazan Genay, Michel Dacorogna, Ulrich Muller, Olivier Pictet, Richard Olsen, *An Introduction to High-Frequency Finance,* Academic Press, 2001

[6] Charles A.E. Goodhart, Maureen O'Hara, *High frequency data in financial markets: Issues and applications,* Journal of Empirical Finance, Volume 4, Issues 23,1997, Pages 73-114

[7] Marcel P. Visser, *GARCH Parameter Estimation Using High-Frequency Data,* Journal of Financial Econometrics, 2011, Vol. 9, No. 1, Pages 162-197.

[8] Siem Jan Koopman, Rutger Lit, Andr Lucas *Intraday Stochastic Volatility in Discrete Price Changes: The Dynamic Skellam Model,* Journal of the American Statistical Association, 2017, Volume 112, Issue 520, Pages 1490-1503.

[9] Alexander Kerss, Nikolai N. Leonenko, Alla Sikorskii, *Fractional Skellam Processes with Applications to Finance,* Fractional Calculus and Applied Analysis, 2014, Volume 17, issue 2, Pages 532-551.

[10] Giovanni De Luca, *Forecasting Volatility Using High-Frequency Data,* Statistica Applicata Vol. 18, n. 2, 2006

[11] Carl Edward Rasmussen, Christopher K. I. Williams, *Gaussian Processes for Machine Learning,* the MIT Press, 2006.

[12] J. Gordon Skellam, *The frequency distribution of the difference between two Poisson variates belonging to different populations,* Journal of the Royal Statistical Society. Series A (General), 1946, 109(Pt 3):296.

CHAPTER 2

Non parametric modeling of Skellam processes. A Gaussian Process approach.

When the likelihood is also a Gaussian, the posterior over the latent function is described by a new GP that is obtained analytically. In all other non-Gaussian cases, e.g. Skellam likelihood, exact inference is intractable and approximate inference methods are needed. In this chapter, we introduce the Laplace approximation and Markov chain Monte Carlo (MCMC) algorithms for inference in GP models.

Without loss of generality, we denote the Skellam process

$$Y_t \sim S(\lambda_{1,t}, \lambda_{2,t}),\ t = 1, 2 \cdots, n$$

where $\lambda_{1,t} = \exp(f_{1,t})$ and $\lambda_{2,t} = \exp(f_{2,t})$. Let $\mathbf{f} = \{\mathbf{f}_1, \mathbf{f}_2\}$ be the latent variables in the Skellam process and θ_i the hyperparameters of $\mathbf{f}_i$.

2.1 The Model

In this Skellam model, we consider the intensities λ_1 and λ_2 are exponential of Gaussian processes. We then summarize the model as follows:

$$\text{Observation model: } \mathbf{Y}|\mathbf{f} \sim \prod_{t=1}^{n} s(y_t; \lambda_{1,t}, \lambda_{2,t}) = \prod_{t=1}^{n} s(y_t; \exp(f_{1,t}), \exp(f_{2,t}))$$

$$\text{GP prior: } f_i(\mathbf{x})|\theta \sim \mathcal{GP}(m_i(\mathbf{x}), k_i(\mathbf{x}, \mathbf{x}'|\theta_i))$$

$$\text{hyperprior: } \theta_i \sim p(\theta_i).$$

Given $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}$, hyperparameter θ_i and the covariance function $k(\mathbf{x}, \mathbf{x}'|\theta_i)$, the GP prior over function $f_i(\mathbf{x})$ have a multivariate Gaussian distribution

$$p(\mathbf{f}_i|\mathbf{X}, \theta_i) = N(\mathbf{f}_i|\mathbf{0}, \mathbf{K}_{\mathbf{f}_i,\mathbf{f}_i})$$

where the covariance matrix $\mathbf{K}_{\mathbf{f}_i,\mathbf{f}_i} = (k(\mathbf{x}_s, \mathbf{x}_t)|\theta_i)_{n\times n}$ is the covariance matrix. Assume that we want to predict the values $\tilde{\mathbf{f}}_i$ at new input locations $\tilde{\mathbf{X}}$. The joint prior for latent variables $\mathbf{f}_i$ and $\tilde{\mathbf{f}}_i$ is

$$\begin{bmatrix}\mathbf{f}_i \\ \tilde{\mathbf{f}}_i\end{bmatrix} |\mathbf{X}, \tilde{\mathbf{X}}, \theta_i \sim N(\mathbf{0}, \begin{bmatrix}\mathbf{K}_{\mathbf{f}_i,\mathbf{f}_i} & \mathbf{K}_{\mathbf{f}_i,\tilde{\mathbf{f}}_i} \\ \mathbf{K}_{\tilde{\mathbf{f}}_i,\mathbf{f}_i} & \mathbf{K}_{\tilde{\mathbf{f}}_i,\tilde{\mathbf{f}}_i}\end{bmatrix}).$$

Therefore, the conditional distribution of $\tilde{\mathbf{f}}_i$ given $\mathbf{f}$ is

$$\tilde{\mathbf{f}}_i|\mathbf{f}_i, \mathbf{X}, \tilde{\mathbf{X}}, \theta_i \sim N(\mathbf{K}_{\tilde{\mathbf{f}}_i,\mathbf{f}}\mathbf{K}^{-1}_{\mathbf{f}_i,\mathbf{f}_i}\mathbf{f}, \mathbf{K}_{\tilde{\mathbf{f}}_i,\tilde{\mathbf{f}}_i} - \mathbf{K}_{\tilde{\mathbf{f}}_i\mathbf{f}_i}\mathbf{K}^{-1}_{\mathbf{f}_i,\mathbf{f}_i}\mathbf{K}_{\mathbf{f}_i,\tilde{\mathbf{f}}_i}).$$

The posterior distribution contains all information about the latent function delivered from the data $\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$,

$$p(\mathbf{f}|\mathcal{D}, \theta) = \frac{p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\mathbf{X}, \theta)}{\int p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\mathbf{X}, \theta)d\mathbf{f}}.$$

Given the conditional posterior distribution $p(\mathbf{f}|\mathcal{D}, \theta)$, we can evaluate the posterior predictive distribution from the conditional mean $E_{\tilde{\mathbf{f}}|\mathbf{f},\theta}[f(\tilde{\mathbf{x}})] = k(\tilde{\mathbf{x}}, \mathbf{X})\mathbf{K}^{-1}_{\mathbf{f},\mathbf{f}}\mathbf{f}$ and the posterior predictive covariance between any set of latent variables, $\tilde{\mathbf{f}}$ is

$$\mathrm{Cov}_{\tilde{\mathbf{f}}|\mathcal{D},\theta}[\tilde{\mathbf{f}}] = E_{\mathbf{f}|\mathcal{D},\theta}[\mathrm{Cov}_{\tilde{\mathbf{f}}|\mathbf{f}}[\tilde{\mathbf{f}}]] + \mathrm{Cov}_{\mathbf{f}|\mathcal{D},\theta}[E_{\tilde{\mathbf{f}}|\mathbf{f}}[\tilde{\mathbf{f}}]].$$

The posterior predictive distribution is

$$\tilde{\mathbf{f}}|\mathbf{f}, \mathcal{D}, \theta \sim \mathcal{GP}(m_p(\tilde{\mathbf{x}}), k_p(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}'))$$

where

$$m_p(\tilde{\mathbf{x}}) = k(\tilde{\mathbf{x}}, \mathbf{X}|\theta)\mathbf{K}^{-1}_{\mathbf{f},\mathbf{f}}E_{\mathbf{f}|\mathcal{D},\theta}[\mathbf{f}]$$

and

$$k_p(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}'|\mathcal{D}, \theta) = k(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}'|\theta) - k(\tilde{\mathbf{x}}, \mathbf{X}|\theta)(\mathbf{K}^{-1}_{\mathbf{f},\mathbf{f}} - \mathbf{K}^{-1}_{\mathbf{f},\mathbf{f}}\mathrm{Cov}_{\mathbf{f}|\mathcal{D},\theta}[\mathbf{f}]\mathbf{K}^{-1}_{\mathbf{f},\mathbf{f}})k(\mathbf{X}, \tilde{\mathbf{x}}'|\theta).$$

To approximate the predictive posterior distribution, we will introduce Laplace approximation and MCMC in the next 2 sections.

2.2 Laplace approximation

Laplace approximation can compute very accurate approximations to the posterior marginals. Rue et al. (2009)[2] showed that the approximations provide more precise estimates in much shorter time. The other advantage of this approach is its generality, which makes it possible to perform Bayesian analysis in an automatic, streamlined way. The Laplace approximation is constructed form the second order Taylor expansion of $\log p(\mathbf{f}|\mathbf{y}, \theta)$ around the mode $\hat{\mathbf{f}}$, which gives a Gaussian approximation to the conditional posterior

$$p(\mathbf{f}|\mathcal{D}, \theta) \sim N(\mathbf{f}|\hat{\mathbf{f}}, \Sigma)$$

where $\hat{\mathbf{f}} = \arg\max_{\mathbf{f}} p(\mathbf{f}|\mathcal{D}, \theta)$ and Σ^{-1} is the Hessian of the negative log conditional posterior at the mode:

$$\Sigma^{-1} = -\nabla\nabla \log p(\mathbf{f}|\mathcal{D}, \theta)|_{\mathbf{f}=\hat{\mathbf{f}}} = \mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1} + \mathbf{W}$$

where $\mathbf{W}$ is a diagonal matrix with entries $\mathbf{W}_{i,i} = \nabla_{f_i}\nabla_{f_i} \log p(y|f_i)|_{f_i=\hat{f}_i}$.

Using $\mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1}\hat{\mathbf{f}} = \nabla \log p(\mathbf{y}|\mathbf{f})|_{\mathbf{f}=\hat{\mathbf{f}}}$, the approximate posterior predictive distribution is given by

$$\tilde{\mathbf{f}}|\mathcal{D}, \theta \sim \mathcal{GP}(m_p(\tilde{\mathbf{x}}), k_p(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}'))$$

where

$$m_p(\tilde{\mathbf{x}}) = k(\tilde{\mathbf{x}}, \mathbf{X})\nabla \log p(\mathbf{y}|\mathbf{f})_{\mathbf{f}=\hat{\mathbf{f}}}$$

and

$$k_p(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}') = k(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}') - k(\tilde{\mathbf{x}}, \mathbf{X})(\mathbf{K}_{\mathbf{f},\mathbf{f}} + \mathbf{W})^{-1}k(\mathbf{X}, \tilde{\mathbf{x}})).$$

The approximate conditional predictive density of $\tilde{y}_i$ can now be evaluated with quadrature integration over each $\tilde{f}_{1,i}$ and $\tilde{f}_{2,i}$ separately.

2.3 Markov Chain Monte Carlo

In MCMC methods, one constructs a Markov chain whose stationary distribution is the posterior distribution and uses the Markov chain samples to obtain Monte Carlo estimates. An advantage of MCMC over deterministic approximate inference is that it provides an arbitrarily precise approximation to the posterior distribution in the limit of long runs. We now follow Titsias et al. (2011)[1] to construct the MCMC algorithm and the sampling strategies.

Let $\theta = \{\theta_1, \theta_2\}$ be the set of parameters in the Gaussian prior $\mathbf{f} = \{\mathbf{f}_1, \mathbf{f}_2\}$. We first sample from the conditional posterior distribution

$$p(\mathbf{f}^{(j+1)}|\mathcal{D}, \theta^{(j)}) \propto p(\mathbf{y}|\mathbf{f}^{(j)})p(\mathbf{f}^{(j)}|\mathbf{X}, \theta^{(j)}) \propto S(\exp \mathbf{f}_1^{(j)}, \exp \mathbf{f}_2^{(j)}) \cdot N(\mathbf{0}, \mathbf{K}_{\mathbf{f}^{(j)},\mathbf{f}^{(j)}}).$$

Next, we sample the hyperparameters in the latent variables (GPs). Sampling from such distributions is carried out by using some proposal distribution, for instance a log uniform.

$$p(\theta^{(j+1)}|\mathbf{f}^{(j+1)}) \propto p(\mathbf{f}^{(j+1)}|\theta^{(j)})p(\theta^{(j)}) \propto N(\mathbf{0}, \mathbf{K}_{\mathbf{f}^{(j+1)},\mathbf{f}^{(j+1)}})p(\theta^{(j)}).$$

After having the posterior sample of latent variables, we can sample from the posterior predictive distribution of any set of $\tilde{\mathbf{f}}$ by sample with each $\mathbf{f}^{(j)}$ from $p(\tilde{\mathbf{f}}|\mathbf{f}^{(j)}, \theta)$. We can obtain a sample of $\tilde{\mathbf{y}}$ from

$$p(\tilde{\mathbf{y}}|\tilde{\mathbf{f}}, \theta) \sim S(\exp \tilde{\mathbf{f}}_1, \exp \tilde{\mathbf{f}}_2).$$

List of References

[1] Michalis K. Titsias, Magnus Rattray, Neil D. Lawrence, *Markov chain Monte Carlo algorithms for Gaussian processes,* in Bayesian Time Series Models, edited by D. Barber, A. T. Cemgil, and S. Chiappa (Cambridge University Press, Cambridge, England), 2011, Pages 295-316.

[2] Hvard Rue, Sara Martino, Nicolas Chopin, *Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations,* Journal of the Royal statistical Society B, 71(2), 2009, pages 1-35.

CHAPTER 3

Simulation Study

In this chapter, we generate two datasets to evaluate this Skellam model and compare with Gaussian model. We first use different covariance functions of Gaussian processes to simulate the intensities in the Skellam process. Next, we will use Laplace approximation and MCMC to show the performance in terms of prediction. When fitting the model, we will utilize GPStuff [1], a Matlab package, which is developed by BECS Bayes group, Aalto University. GPstuff is a versatile collection of Gaussian process models and computational tools required for inference. This toolbox is designed to allow adding new model blocks in the package, e.g. likelihood, covariance function, and prior. Therefore, we add Skellam likelihood into this package and use Laplace approximation and MCMC.

3.1 Dataset I

Consider the most widely-used covariance function, the squared exponential

$$k(x, x') = \sigma^2 \exp(-\frac{||x - x'||^2}{2 \cdot l^2}).$$

The generating process is describe as below: we select different values for the length scale l and the magnitude σ^2 to generate GPs. Let

$$k_1(x, x') = \sigma_1^2 \exp(-\frac{||x - x'||^2}{2 \cdot l_1^2}) = 0.5 \cdot \exp(-\frac{||x - x'||^2}{2 \cdot 0.45^2})$$

and

$$k_2(x, x') = \sigma_2^2 \exp(-\frac{||x - x'||^2}{2 \cdot l_2^2}) = \exp(-\frac{||x - x'||^2}{2 \cdot 1.73^2}).$$

$\mathbf{f}_1$ and $\mathbf{f}_2$ are Gaussian process with covariance functions k_1 and k_2, respectively. We now let the Skellam model

$$Y \sim \text{Poisson}(\lambda_1) - \text{Poisson}(\lambda_2)$$

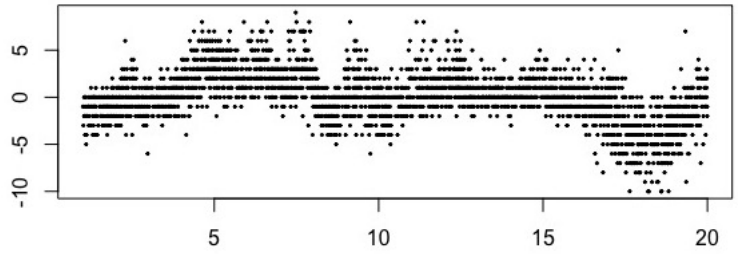


Figure 1: Simulated Data I: 3000 observations

where the intensities $\lambda_i = \exp(\mathbf{f}_i + N(0, \sigma^2))$ are the GPs with noise $\sigma = 0.1$. Figure 1 is the generated dataset.

3.1.1 Laplace approximation

We now fit this dataset into the Skellam model via Laplace approximation. We simply assign log uniform distribution to all hyperparameters of the Gaussian priors. Then the Laplace approximation to the posterior distribution which can estimate the latent values and prediction of $\mathbf{y}$. We will discuss the comparisons between the actual values of the latent variables and $\mathbf{y}$ below.

Recall that latent values are generated via Gaussian process. The intensities in the Poisson distributions are created by adding noise to the latent values. Figure 2 compares the latent values between true values and predicted values. In Figure 3, we can see that the prediction of Skellam model is well centered in the noisy true values.

Comparison of the difference of Poisson distributions and the difference of the predicted latent values, $E\lambda_1 - E\lambda_2$, are shown in Figure 4. In this figure, we also find the comparison between $\mathbf{y}$ and difference of Poisson processes.

3.1.2 MCMC approach

We now use MCMC to approximate the model. Although MCMC is a powerful algorithm in Bayesian methods, sampling posterior with MCMC is still time-

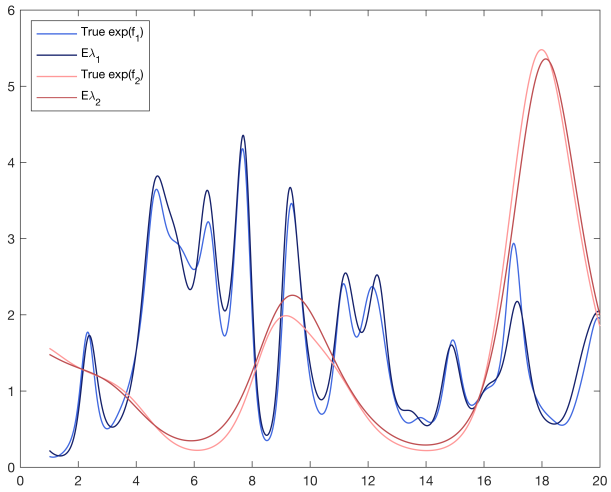


Figure 2: Comparison of true and predicted latent values

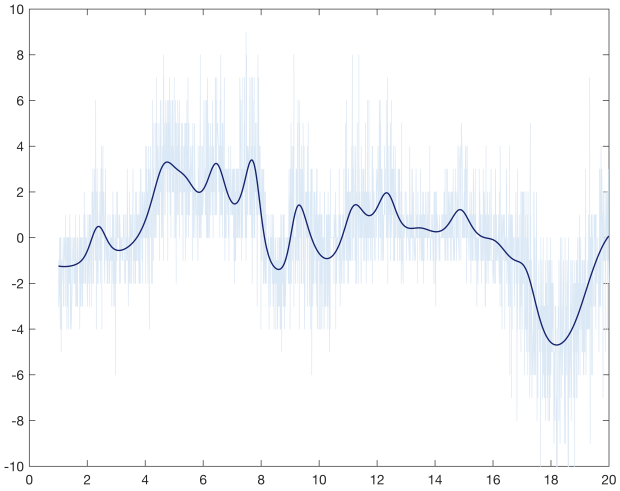


Figure 3: Prediction of the Skellam models via Laplace approximation

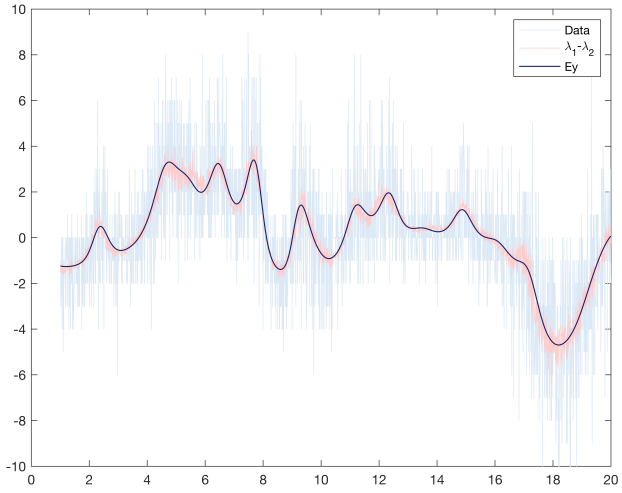


Figure 4: Comparison of the difference of Poisson distributions and the difference of the predicted latent values (Laplace)

consuming. Therefore, we use less observations, 600 observations, to process the MCMC. Unlike the Laplace approximation, assigning a log uniform prior to the logarithm of the hyperparameters of the covariance function lead to inefficient sampling. Therefore I used a $N(0, 2)$.

Sampling multiple hyperparameters can be a complex process. It sometimes is difficult to converge well and hard to predict well on all hyperparameters. Figure 5 shows the trace plots of hyperparameters of the Gaussian priors and Figure 6 are the posterior distributions of each hyperparameters. Recall that, in Figure 5, σ_i^2 and l_i correspond to the hyperparameters in the covariance functions

$$k_1(x, x') = \sigma_1^2 \exp(-\frac{||x - x'||^2}{2 \cdot l_1^2}) \text{ and } k_2(x, x') = \sigma_2^2 \exp(-\frac{||x - x'||^2}{2 \cdot l_2^2}).$$

Table 1 present the mean and 95% credible interval (CI) in the sampled hyperparameters. It tells the true values of signal variance, σ_1^2.

The MCMC algorithm seems did not recover the values of hyperparameters well. However, in Figure 7, if we compare the true latent values and predicted latent values, one can see that the trends of latent values are moderately recovered.

	True Value	mean	95% CI
σ_1^2	0.5	1.7057	(0.6242, 3.4795)
l_1	0.45	0.5977	(0.3667, 0.8619)
σ_2^2	1	0.4303	(0.0265, 1.9220)
l_2	1.73	1.875	(0.4150, 3.7924)

Table 1: Summary statistics of the hyperparameters of the posterior distribution: Simulated Data I

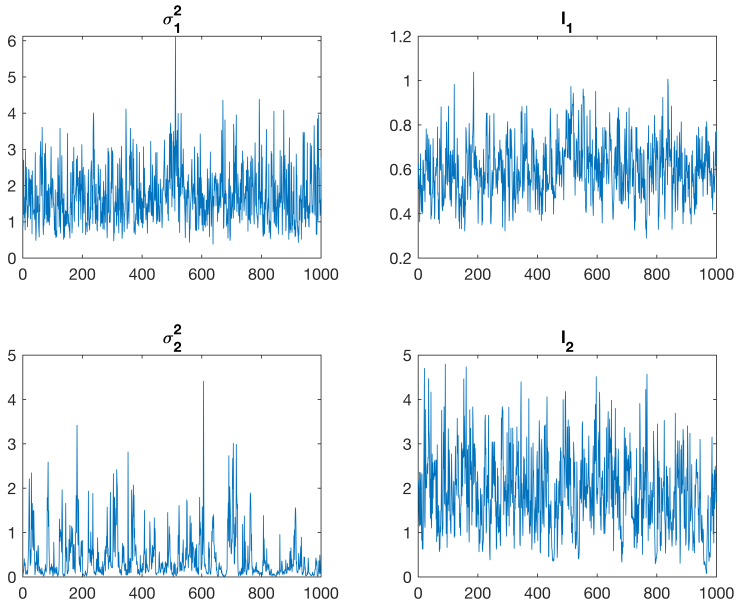


Figure 5: Trace plots

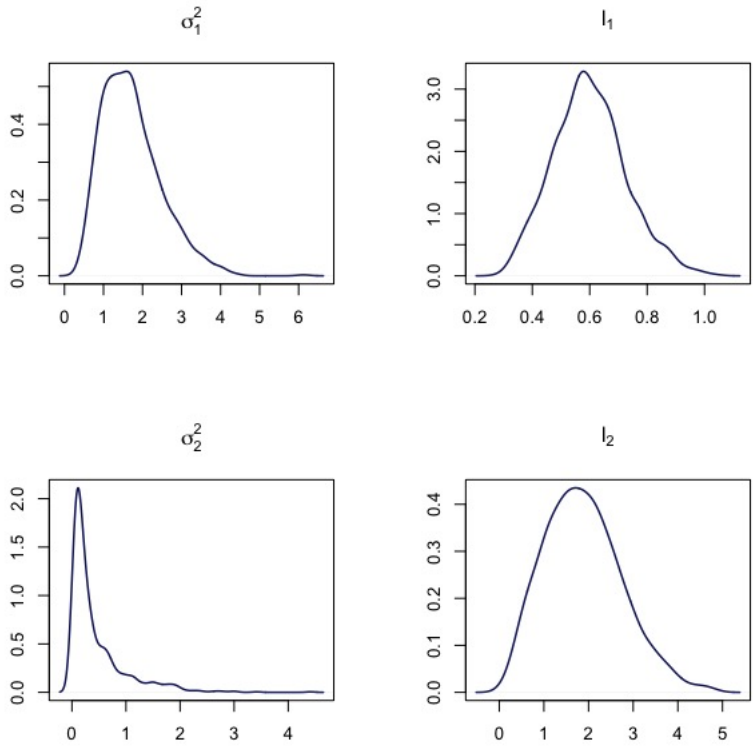


Figure 6: Posterior kernel density estimates

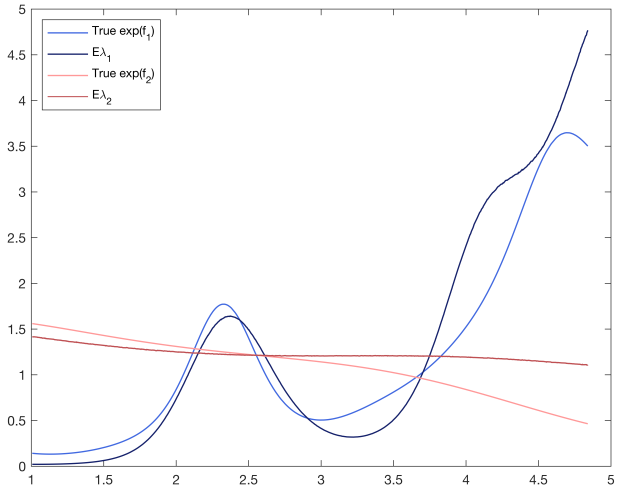


Figure 7: The comparison of true and predicted latent values via MCMC approach

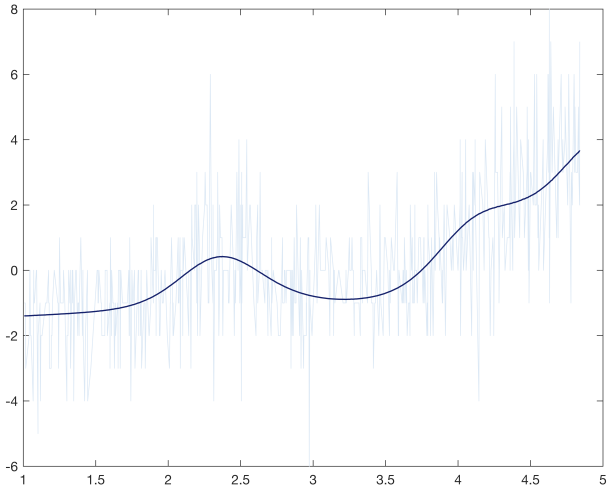


Figure 8: The comparison of true and predicted Skellam models (MCMC)

Figure 8 gives predictions on MCMC approaches. Although MCMC might not well recover the latent values and hyperparameters, it still well predicted the trends of the intensities of two Poisson process in the Skellam process.

3.1.3 Dataset II

We now consider one of the covariance function to be the periodic covariance function

$$k_(x, x') = \sigma^2 \exp(-\frac{2\sin^2(\pi(x - x')/\gamma)}{l^2})$$

where the parameter γ controls the inverse length of the periodicity and l the smoothness of the process. We simulate the data with one squared exponential and one periodic covariance functions

$$k_1(x, x) = 0.5 \cdot \exp(-\frac{||x - x'||^2}{2 \cdot 0.45^2})$$

$$k_2(x, x') = 1 \cdot \exp(-\frac{2\sin^2(\pi(x - x')/4)}{3^2}).$$

Again, the Skellam model is

$$Y \sim \text{Poisson}(\lambda_1) - \text{Poisson}(\lambda_2)$$

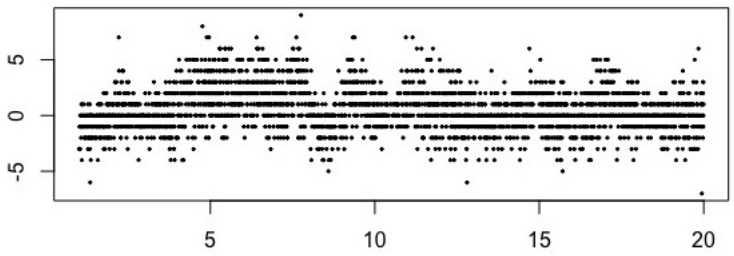


Figure 9: Simulated Data II: 3000 observations

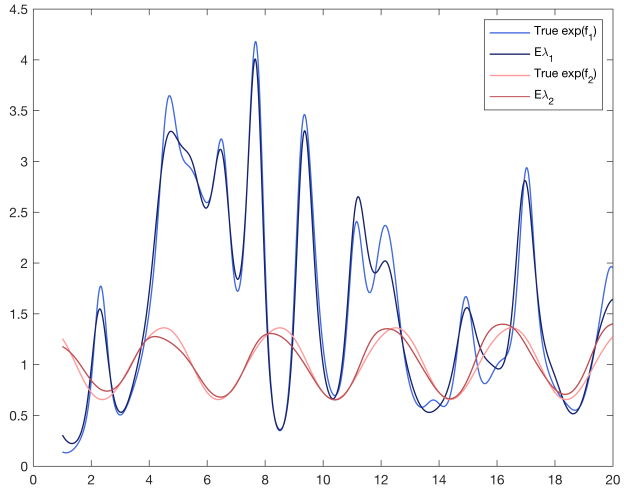


Figure 10: Comparison of true and predicted latent values (Laplace)

where $\lambda_i = \exp(\mathbf{f}_i + N(0, \sigma^2))$ and $\mathbf{f}_i$ is GP with covariance function k_i.

The generated data is shown in Figure 9.

3.1.4 Laplace approximation

The comparison of true and predicted latent values is in Figure 10, and the prediction of Skellam model, $\hat{\mathbf{y}}$ can be seen in Figure 11. Comparison of the difference of Poisson distributions and the difference of the predicted latent values, $E\lambda_1 - E\lambda_2$, are shown in Figure 12. In this figure, we will again find the comparison between $\mathbf{y}$ and difference of Poisson processes.

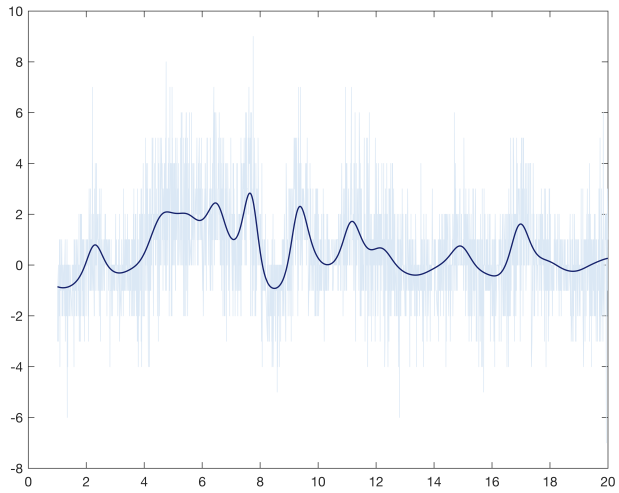


Figure 11: Prediction of the Skellam models (Laplace)

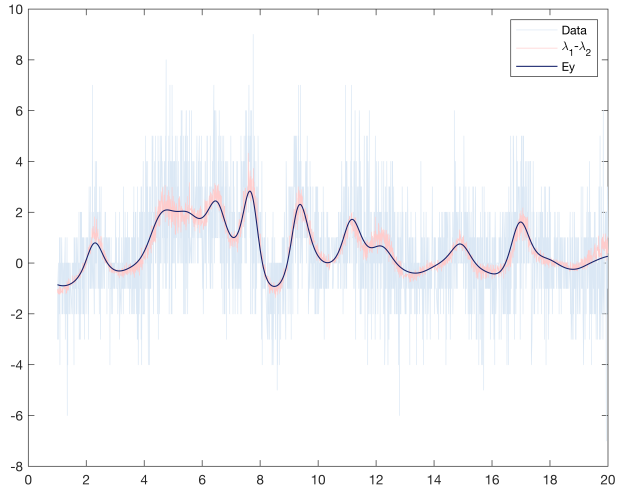


Figure 12: Comparison of the difference of Poisson distributions and the difference of the predicted latent values (Laplace)

	True Value	mean	95% CI
σ_1^2	0.5	1.7156	(0.6457, 3.4064)
l_1	0.45	0.6200	(0.3849, 0.8557)
σ_2^2	1	0.5370	(0.0203, 2.0931)
l_2	3	1.4430	(0.3454, 3.1813)

Table 2: Summary statistics of the hyperparameters of the posterior distribution: Simulated Data II

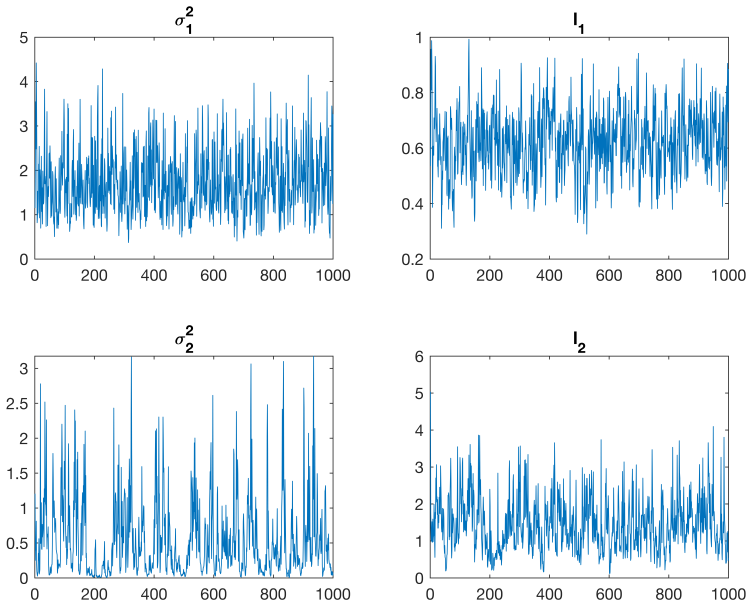


Figure 13: Trace plots

3.1.5 MCMC approach

Figure 13 show the trace plots of hyperparameters of the Gaussian priors and Figure 14 are the posterior distributions of each hyperparameters. Recall that, in Figure 13, σ_i^2 and l_i correspond to the hyperparameters in the covariance functions

$$k_1(x,x) = \sigma_1^2 \cdot \exp(-\frac{||x-x'||^2}{2 \cdot l_1^2}) \text{ and } k_2(x,x') = \sigma_2^2 \cdot \exp(-\frac{2\sin^2(\pi(x-x')/4)}{l_2^2}).$$

Table 2 present the mean and 95% CI in the sampled hyperparameters. Figure 15 compares the true latent values and predicted latent values. Figure 16 gives predictions of the second dataset.

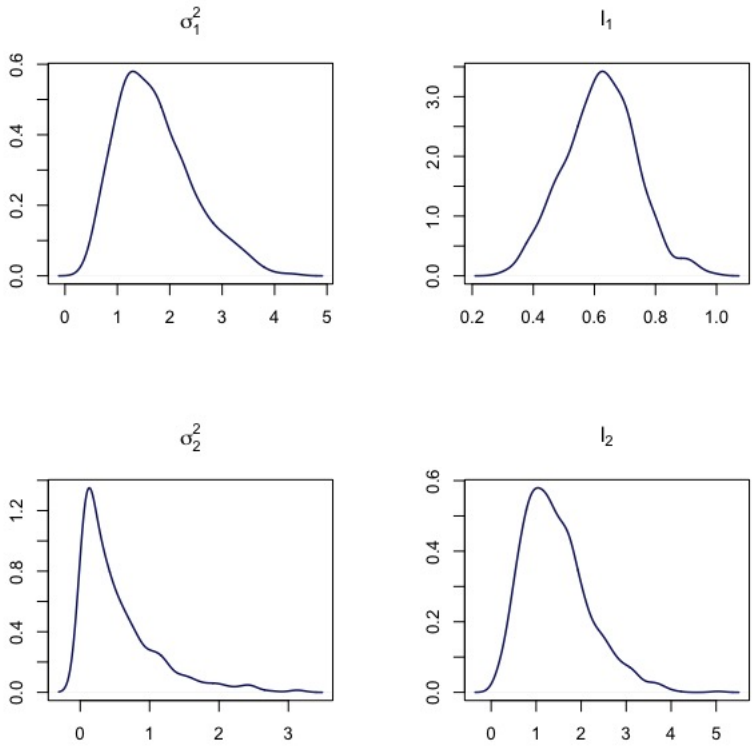


Figure 14: Posterior kernel density estimates

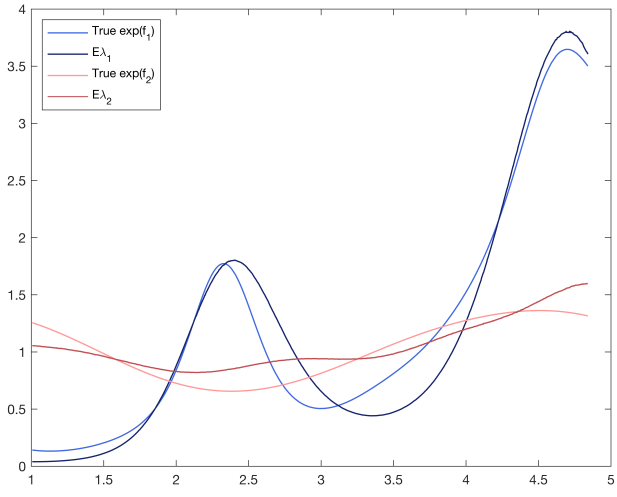


Figure 15: The comparison of true and predicted latent values (MCMC)

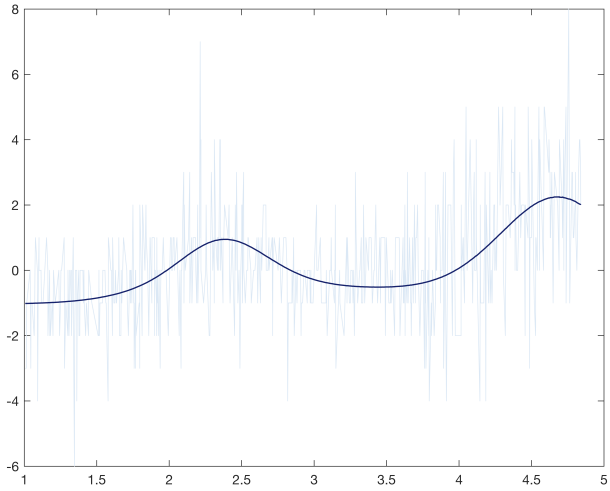


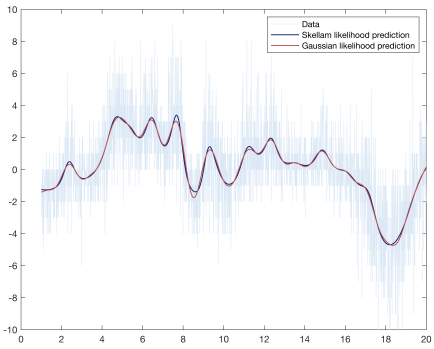
Figure 16: The comparison of true and predicted Skellam models(MCMC)

	Skellam likelihood	Gaussian likelihood
Simulated Data I	-1.7995	-1.8579
Simulated Data II	-1.8615	-1.8551

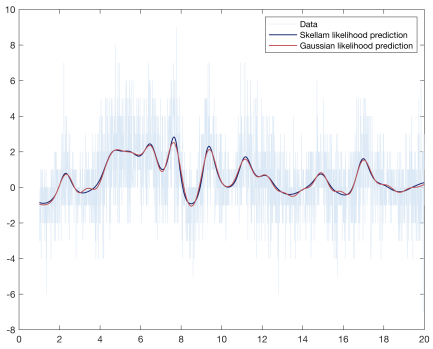
Table 3: DIC

3.2 Skellam likelihood and Gaussian likelihood

According to the problems that we are interested in, we fit the data to the model that suitable to our purpose. In this section, we fit the datasets into a single latent variable model, Gaussian likelihood. Gaussian likelihood might be cheaper in computation, but it obvious can not model the intensities in the Poisson processes. Therefore, we use Gaussian likelihood only to compare the predictive distribution. Figure 17 and 18 compares the predictions via Laplace approximation and MCMC. Skellam and Gaussian likelihoods have similar performance in Laplace approximation and MCMC. The deviance information criterion (DIC) in table 3 also gives the same conclusion. .

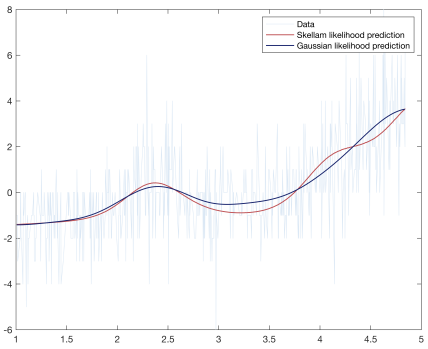


(a) Simulated Data I

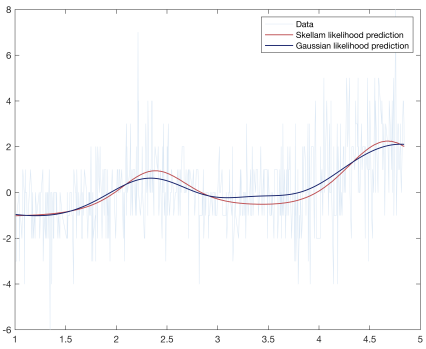


(b) Simulated Data II

Figure 17: Skellam and Gaussian likelihoods (Laplace)



(a) Simulated Data I



(b) Simulated Data II

Figure 18: Skellam and Gaussian likelihoods (MCMC)

List of References

[1] Jarno Vanhatalo, Jaakko Riihimki, Jouni Hartikainen, Pasi Jylnki, Ville Tolvanen, and Aki Vehtari, *GPstuff: Bayesian Modeling with Gaussian Processes,* Journal of Machine Learning Research, 2013, 14(Apr):1175-1179.

CHAPTER 4

Application to high frequency financial data

The datasets used in this chapter are the price changes in 1-min frequency from S&P 500 index, NASDAQ, and IBM stock trading data. S&P 500 index and NASDAQ are futures contract data. A futures contract is a legal agreement to buy or sell a particular commodity or asset at a predetermined price at a specified time in the future. Futures contracts are standardized for quality and quantity to facilitate trading on a futures exchange. The buyer of a futures contract is taking on the obligation to buy the underlying asset when the futures contract expires. The seller of the futures contract is taking on the obligation to provide the underlying asset at the expiration date.

We will discuss the model fitting for each datasets below. We use 5000 observations to fit the model via Laplace approximation and 1000 observations via MCMC approach. In this chapter, we assign Gaussian priors with squared exponential covariance functions. That is,

$$Y \sim S(\exp(\mathbf{f}_1), \exp(\mathbf{f}_2))$$

where $f_i(\mathbf{x})|\theta_i \sim \mathcal{GP}(m_i(\mathbf{x}), k_i(\mathbf{x}, \mathbf{x}'|\theta_i))$ and $k_i(\mathbf{x}, \mathbf{x}') = \sigma_i^2 \exp(\frac{||\mathbf{x} - \mathbf{x}'||^2}{2 \cdot l_i^2})$.

4.1 S&P 500 index

The time period for this dataset is 2019/06/17 09:01 to 2019/06/21 01:33 in 1-minute scale. Figure 19 gives the price change during this time period of SP 500. The prediction via Laplace approximation of price changes and latent values are shown in Figure 20. Figure 21, 22 and 23 are the trace plots and the predictions via MCMC approach.

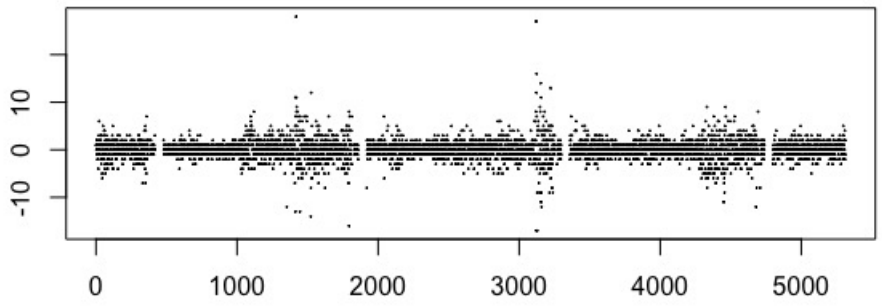


Figure 19: SP 500: Price changes on 2019/06/17 to 2019/06/21

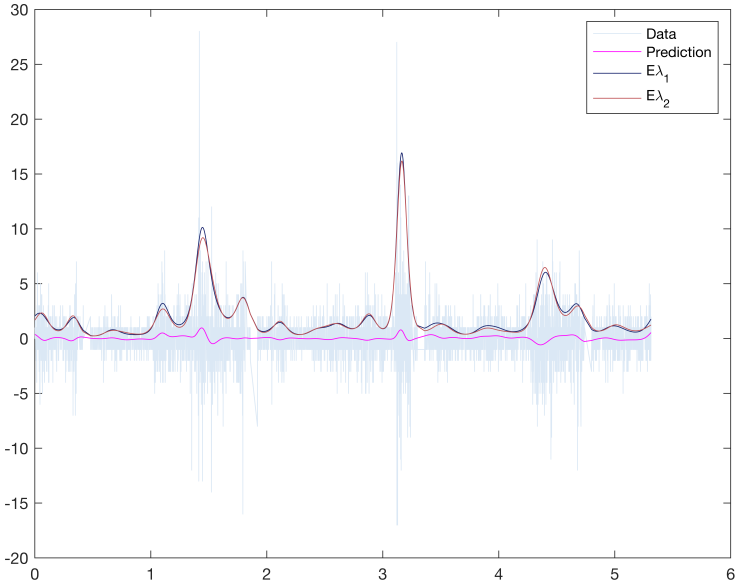


Figure 20: The prediction of SP500 via Laplace approximation

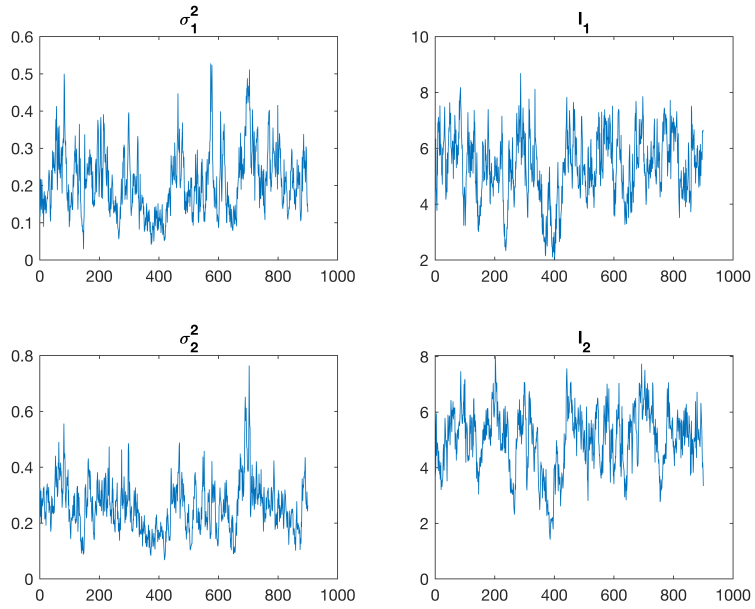


Figure 21: SP500: Trace plots

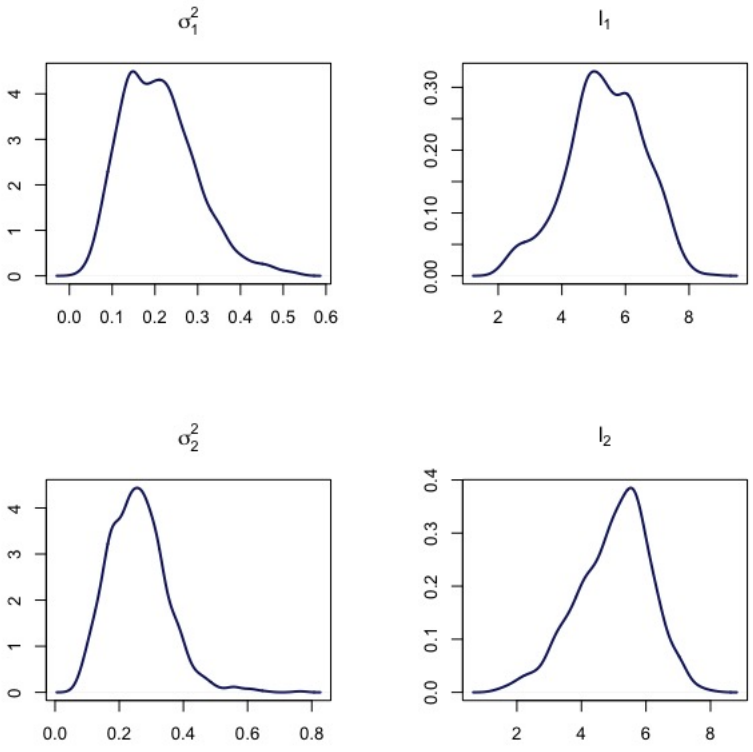


Figure 22: SP 500: posterior kernel density estimates

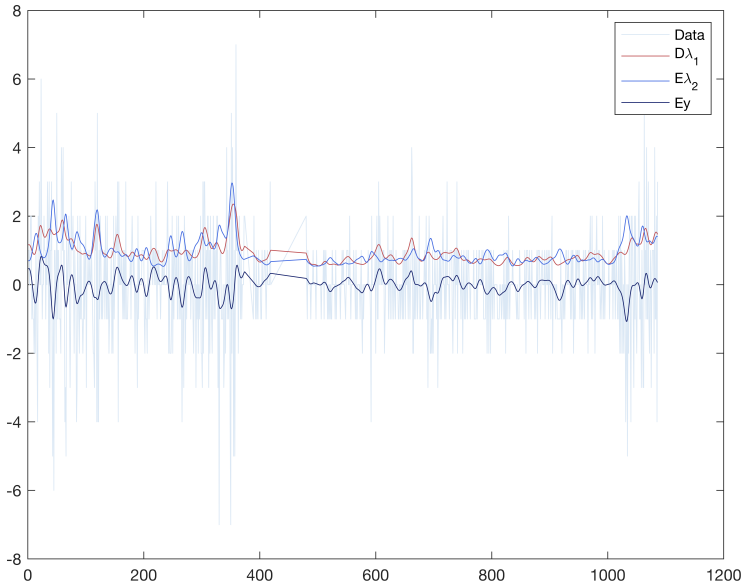


Figure 23: The prediction of SP500 via MCMC approach

4.2 NASDAQ

The time period of NASDAQ dataset is also 2019/06/17 09:01 to 2019/06/21 01:33 in 1-minute scale. Unfortunately, Laplace approximation is not working on this dataset due to computational stability reason. We can see that the price change rapidly in Figure 24. Therefore, we will present the result of MCMC approach here. Figure 25 and 26 are the trace plots and the posterior distributions of the hyperparameters. The predictions of the Skellam model and latent values are in Figure 27.

4.3 IBM Trading

For the IBM trading data, the time period is 2019/06/10 8:31 to 2019/06/26 13:27 in 1-minute scale. Unlike futures contract financial data, the market opens at 8:00 and closes at 15:15 everyday. There are gaps of the data between days (see Figure 28). In Figure 29, we bridge the gaps to fit into the model.

While fitting the data into Skellam model via Laplace approximation, we

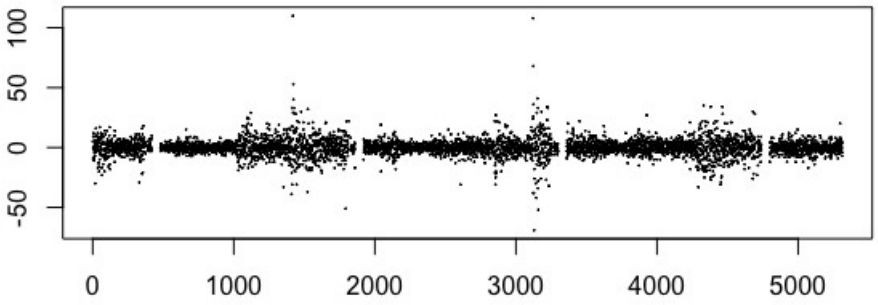


Figure 24: NASDAQ: Price changes on 2019/06/17 to 2019/06/21

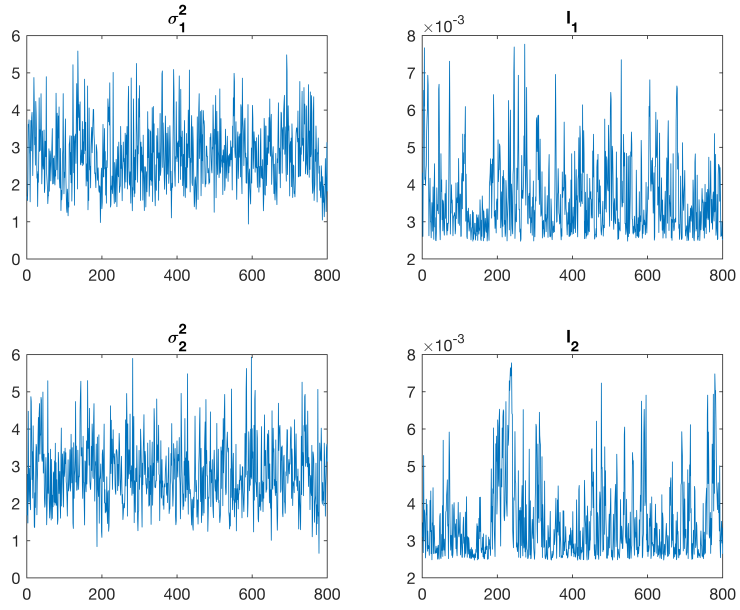


Figure 25: NASDAQ: Trace plots

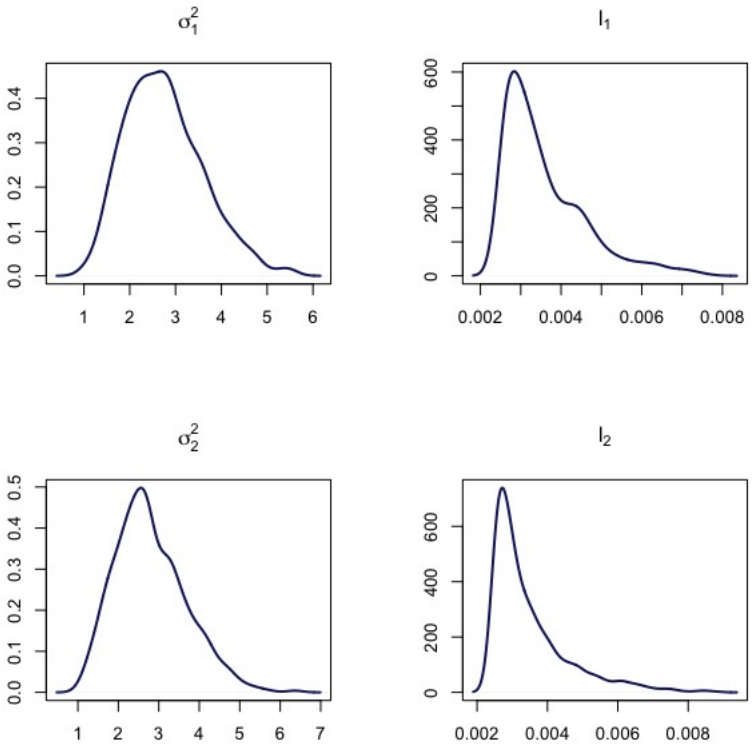


Figure 26: NASDAQ: posterior kernel density estimates

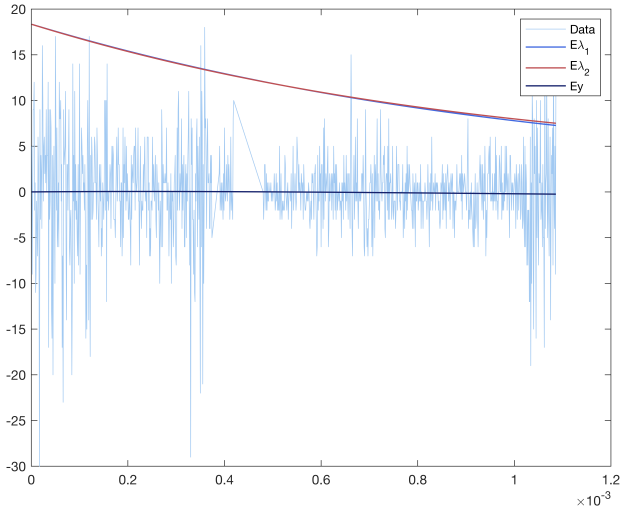


Figure 27: NASDAQ: MCMC predictions

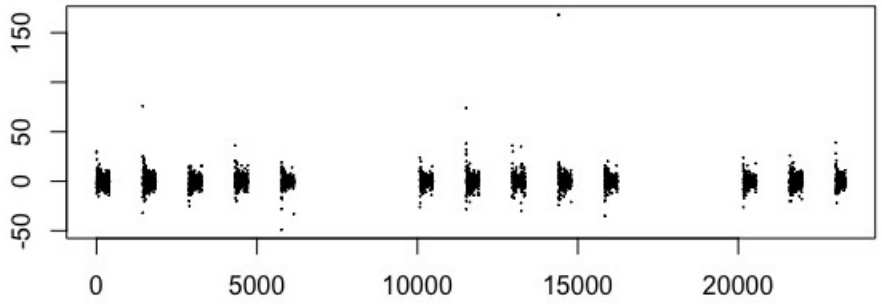


Figure 28: IBM: Price changes on 2019/06/10 8:31 to 2019/06/26 13:27

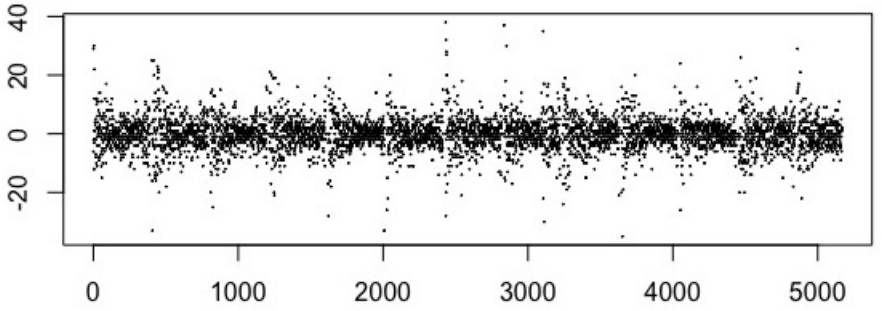


Figure 29: IBM: Price changes on 2019/06/10 8:31 to 2019/06/26 13:27 after bridging

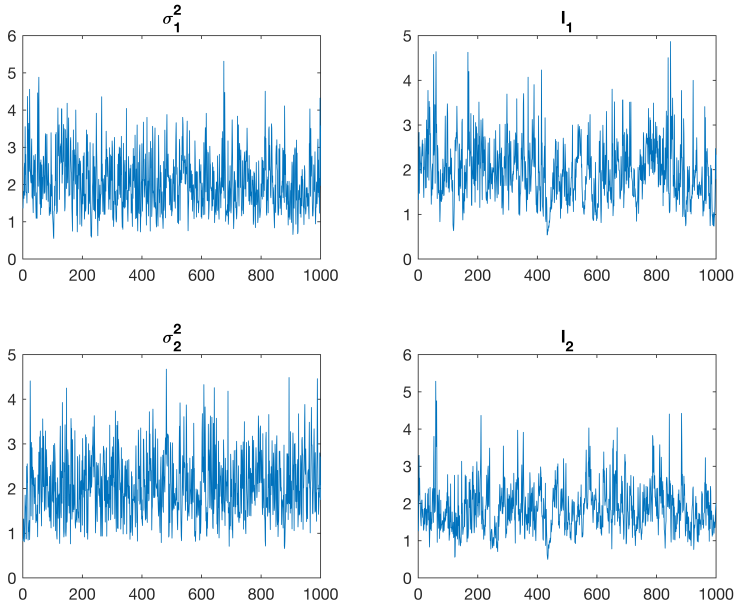


Figure 30: IBM: Trace plots

face the same issue as NASDAQ. Therefore, we will show the result with MCMC approach. Figure 30 are the trace plots of the MCMC sampling and Figure 31 are the posterior distributions of the hyperparameters. Figure 32 is the predictions of the Skellam model and latent values.

4.4 Discussion

These three datasets present relatively quiet market. Figure 23, 27, and 32 show that the predictions of latent values are almost overlapping to each other. While fitting the three datasets into Skellam model via Laplace approximation, we faced computational stability challenges in NASDAQ and IBM datasets. It might be caused by the larger range of price changes in NASDAQ and IBM. We need further study to understand the issues.

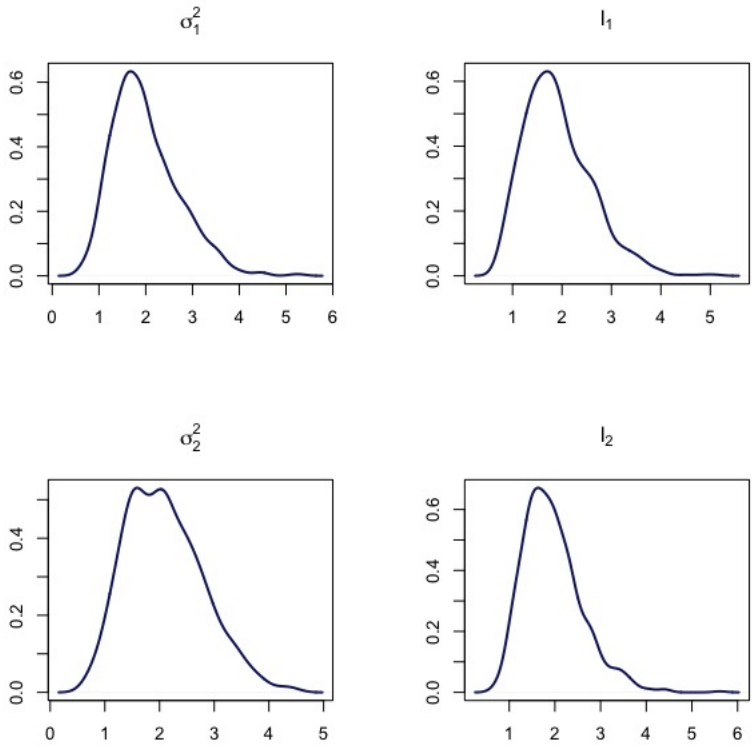


Figure 31: IBM: posterior kernel density estimates

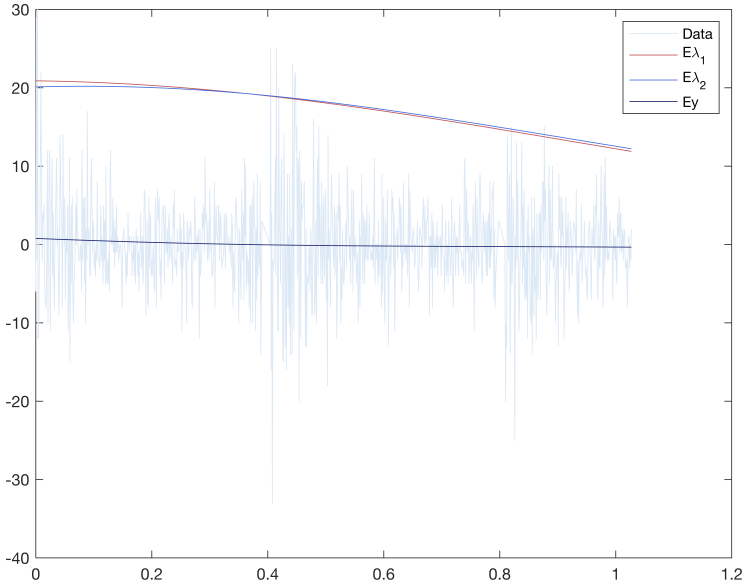


Figure 32: The prediction of IBM via MCMC approach

CHAPTER 5

Discussion and future research

Recent academic research has focus on valuation techniques to uncover the market microstructure such as trading and market structure, market rules and fairness, and price discovery, etc, in high frequency financial data. In modeling such data, Skellam process provides not only the price change prediction, also captures the latent variables that are correlated to the transaction volume, supply and demand, in a tick time.

Laplace approximation and Markov chain Monte Carlo approaches have different advantages. Laplace approximation offers the efficiency in predicting the price change and unfolding the latent variables. Laplace approximation has limitation in time scale. Inappropriate time scale leads to computational burden. Thus, sometimes the time variable needs to be rescaled. However, we don't always find a good time scale for a specific dataset. MCMC provides more flexibility in data selection and capacity in obtaining the estimations of hyperparameters in the latent variables.

In our simulation study, the two algorithms perform very well in estimating the latent trajectories of the intensity functions While we fitting the Skellam model into S&P 500, NASDAQ, and IBM dataset via Laplace approximation approach, the time scale was sensitive. We had to adjust the scale on time variable. However, we haven't find the general rule in scaling the time variable.

Future research directions:

1. Improve the computational efficiency of the Skellam process due to the high volume of data.

2. Discover the subtlety of time variable among the indices and stocks datasets.

3. Apply Skellam model to other types of data with Skellam distributed counts.

4. Investigate the Skellam model when the intensities, λ_1 and λ_2, are not independent.

BIBLIOGRAPHY

C. E. Rasmussen, C. K. I. W., *Gaussian Processes for Machine Learning.* Massachusetts Institute of Technology: the MIT Press, 2006.

Gelman, A., *Bayesian Data Analysis, third edition.* Boca Raton: CRC Press, 2014.

Hoff, P. D., *A First Course in Bayesian Statistical Methods.* New York, NY: Springer, 2009.